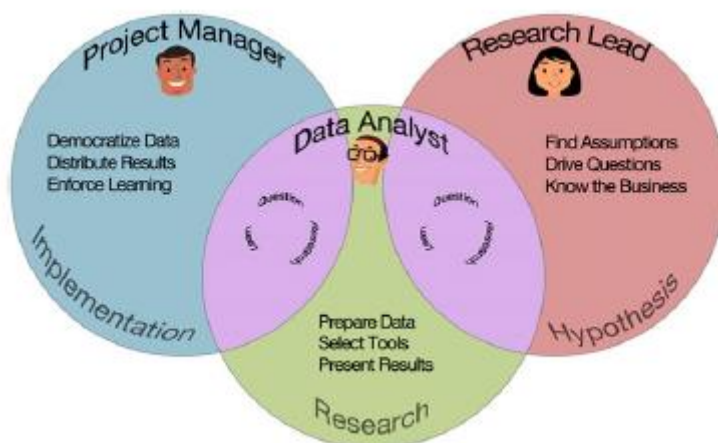


Rozpoczęcie pracy

Teraz, gdy masz już swój zespół, musisz zrozumieć, jak zacząć pracę. Najpierw musisz zbadać obszary odpowiedzialności każdego członka zespołu i wskazać, gdzie te obszary odpowiedzialności się pokrywają. Następną rzeczą, o której musisz pomyśleć, jest sposób prezentacji danych. Wizualizacja danych to temat poruszany w całych książkach. Pokróćce poruszę ten temat i omówię dwa różne typy raportów, na których Twój zespół ds. analityki danych będzie się głównie koncentrował: wewnętrzne i zewnętrzne. Gdy już wiesz, jak zamierzasz prezentować swoje raporty, przejdę do wyjaśnienia silosów danych i ich wpływu na zdolność Twojego zespołu do zbierania danych. Następnie podam kilka wskazówek, jak zdemokratyzować swoje dane. Na koniec omówię znaczenie zapewnienia wglądu w zespół ds. analizy danych i uzyskania poparcia dla kierownictwa.

Definiowanie obszarów odpowiedzialności

Nastawienie na naukę danych to duża zmiana w stosunku do sposobu działania większości organizacji. Nawet organizacje, które nazywają się data-driven, często nie wykorzystują swoich danych do tworzenia nowych spostrzeżeń. Zamiast tego używają danych w taki sam sposób, w jaki pijana osoba korzysta z latarni. Uważają to za wsparcie, a nie iluminację. To prawdziwe wyzwanie dla Twojego zespołu zajmującego się analizą danych. Twoja organizacja może myśleć, że jest oparta na danych, ale w rzeczywistości używa danych tylko do wzmocnienia tego, co już wie. Wszystko, co zaprzecza tej wiedzy, jest postrzegane jako złe dane. Twój zespół ds. analityki danych musi upewnić się, że wykorzystuje dane do odkrywania, dzięki czemu zespół nie wpada w pułapkę polegającą na używaniu danych tylko do wspierania tego, co już jest znane. W rzeczywistości główną korzyścią nauki o danych jest kwestionowanie ustalonej wiedzy. To jak ten stary cytat Marka Twaina: „To, co wpędza nas w kłopoty, nie jest tym, czego nie wiemy. Wiemy na pewno, że tak nie jest. Jeśli Twoja organizacja polega na wiedzy, która nie jest poparta danymi, prawdopodobnie wpadniesz w kłopoty. Często ta wspólna wiedza jest słuszną, ale gdy jest złą, może mieć trwałe konsekwencje. Jeśli Twój zespół zajmujący się badaniem danych pozostaje wierny swoim trzem obszarom obowiązków, może to być realną korzyścią dla Twojej organizacji. Te trzy obszary to badania, pytania i wdrażanie, jak omówiono w kolejnych sekcjach. Do tej pory widziałeś trzy typowe role w zespole zajmującym się badaniem danych: kierownik badania, który zadaje interesujące pytania, analitycy danych, którzy pracują z badaniami prowadzącymi badania w celu opracowania interesujących raportów i spostrzeżeń, oraz kierownik projektu, który sprawia, że te spostrzeżenia są przydatne i dostępne dla reszty organizacji. Teraz nadszedł czas, aby przejść te role i umieścić je w szerszych obszarach odpowiedzialności, aby zobaczyć, jak łączy się zespół. Pomyśl o swoim zespole jako o różnych nakładających się obszarach odpowiedzialności



Badania

Zacznijmy nasz diagram Venna od analityka danych. Podstawowym obszarem odpowiedzialności analityka danych są badania, które są kluczową częścią nauki i nauki o danych. Analityk danych współpracuje z prowadzącym badaniem, aby wymyślić interesujące pytania, zbadać te pytania i przedstawić wyniki za pomocą kreatywnego raportu lub wykresu. Ten analityk danych jest podstawą zespołu. Współpracuje zarówno z kierownikiem projektu, jak i kierownikiem badań. Po prostu działają na różne sposoby. Pomyśl o nim jako o niezależnej dwukierunkowej relacji z nimi obojętnie. Otrzymuje informacje od kierownika badań w formie interesujących pytań, a następnie przekazuje wyniki i spostrzeżenia kierownikowi projektu, aby mogła wymusić uczenie się.

Pytający

Teraz spójrzmy na następny okrąg na rysunku. Po prawej stronie koło prowadzącego badania. Ich obszar odpowiedzialności to kwestionowanie. Jeśli myślisz o tym w kategoriach metody naukowej, to ta osoba tworzy interesującą hipotezę. Prowadząca badania tworzy cykl między sobą a analitykiem danych. Zadaje pytania i otrzymuje informacje zwrotne. Nie jest to tak proste, jak wysłanie e-maila do analityka danych z pytaniem „Co o tym myślisz?” To proces współpracy. Kierownik badania zadaje pytania, a analityk danych udziela informacji zwrotnej na te pytania na podstawie dostępnych danych. Te dwa kręgi nakładają się na siebie. To bezpośredni związek między pytaniami a badaniami.

Realizacja

Ostatnie kółko na rysunku jest przeznaczone dla kierownika projektu. Jego obszarem odpowiedzialności jest wdrażanie. Musi upewnić się, że zespół weźmie dane i wykorzysta je do czegoś, co da się wykonać. Upewnia się, że zespół przekazuje swoje spostrzeżenia reszcie organizacji. Podjęcie procesu eksploracyjnego i zastosowanie go do wiedzy organizacyjnej nie jest łatwym wyzwaniem. W zespole zajmującym się analizą danych często nie wiesz, jakie są Twoje najbardziej przydatne wnioski. Zespół przejdzie przez 50 ślepych uliczek, zanim znajdzie jedną interesującą ścieżkę. Mimo to ważne jest, aby zastanowić się, jak te spostrzeżenia mogą wyglądać, gdy zostaną ostatecznie wdrożone. Każdy z tych obszarów odpowiedzialności jest częścią mapy tego, co zespół musi osiągnąć. Pomaga to wzmocnić ideę, że zespół ds. analizy danych zajmuje się eksploracją i odkrywaniem. Zespół musi śledzić dane, nawet jeśli stoją w sprzeczności z ustaloną wiedzą.

Prezentacja raportów

Wizualizacja danych to jeden z najciekawszych obszarów nauki o danych. W ten sposób wyświetlasz swoje dane za pomocą grafiki i obrazów. Prosta grafika to jeden z najlepszych sposobów komunikowania złożonych pomysłów. Zrównoważenie złożonych pomysłów z prostym projektem może być prawdziwym wyzwaniem. Większość raportów wskazuje zbyt daleko w jednym lub drugim kierunku. Istnieje wiele pięknych wykresów, które tak naprawdę nie przekazują zbyt wielu informacji, są też wykresy informacyjne, które są brzydkie i trudne do zrozumienia. Dobra wizualizacja to podstawowy obowiązek analityka danych. Biorąc to pod uwagę, analityk powinien ściśle współpracować z zespołem ds. analityki danych. Jeśli musisz wyjaśnić wykres kierownikowi badań, prawdopodobnie jest on zbyt skomplikowany dla wszystkich innych. Zespół jest dobrym poligonem doświadczalnym, który pomoże Ci uczynić Twoje wykresy piękniejszymi i czytelnymi. Istnieje wiele dobrych zasobów dotyczących wizualizacji danych. Jednym z najstarszych i najbardziej szanowanych jest Wizualne wyświetlanie informacji ilościowych. W tej książce profesor Tufte przedstawia ideę stosunku danych do atramentu. Twierdzi, że powinieneś przesyłać swoje wizualizacje, aby przekazywać maksymalną ilość danych przy minimalnej ilości atramentu. Używa terminu chartjunk do bezużytecznych wizualizacji, takich jak cienie trójwymiarowe lub efekty gradientowe. Prof. Tufte

pomógł w ustaleniu i wyznaczeniu kierunku dziedziny wizualizacji danych. Pomógł stworzyć nowoczesne podejście do prezentacji danych. Aby zapoznać się z innym podejściem, zapoznaj się z artykułem Opowiadanie historii z danymi: Przewodnik po wizualizacji danych dla profesjonalistów biznesowych autorstwa Cole Nussbaumera Knaflica. W swojej książce przedstawia sześć kluczowych lekcji. Zaczyna od stwierdzenia, że analityk danych powinien rozumieć swoich odbiorców i kontekst; wybierz odpowiedni wyświetlacz, wyeliminuj bałagan i skup uwagę; i myśl jak projektant, aby opowiedzieć fascynującą historię. Te dwie książki dadzą ci dobre wyobrażenie o tym, co to znaczy tworzyć ładne wizualizacje danych.

Uwaga : Zazwyczaj szkolenie analityka danych nie przygotowuje go do tworzenia dobrych wizualizacji. Większość programów podyplomowych jest nadal mocno zakorzeniona w matematyce i statystyce. Dobra wizualizacja danych opiera się na estetyce i designie. To wycuczona umiejętność i może nie być łatwa.

Pamiętaj, że Twój zespół pracuje razem nad eksploracją danych, co oznacza, że większość pierwszej rundy raportów, które zaprojektujesz, będzie dla siebie nawzajem. Kierownik badań kieruje interesującymi pytaniami, analityk danych tworzy szybki i brudny raport, aby zbadać możliwe odpowiedzi, a następnie zespół może wymyślić całą serię nowych pytań. Oznacza to, że większość początkowych wizualizacji danych będzie polegała na szybkiej wymianie — bardziej jak wizualna pogawędka niż pełne raporty danych. Z drugiej strony istnieją ostateczne wizualizacje danych, które musisz komunikować z resztą organizacji. Te wizualizacje będą dopracowane, prostsze i łatwiejsze do zrozumienia. Pomyśl o swojej pierwszej rundzie wizualizacji jak o tablicach z danymi, takich jak tablice na ścianach większości organizacji. Wykorzystywane są głównie do szybkich wizualizacji pomagających w małych spotkaniach i dyskusjach. Twój pierwszy raport powinien mieć taki poziom czytelności. Może to być szybki i prosty diagram klastrowy. Może będzie to po prostu prosta regresja liniowa. Nikt nigdy nie zrobiłby zdjęcia bazgrołów na tablicy zawartej w raporcie wykonawczym. To tylko dla zespołu. Aby przenieść reprezentację danych z tablicy do raportu wykonawczego, musisz dodać przejrzystości i dopracowania. Jeśli jesteś analitykiem danych, pamiętaj, aby skorzystać z opinii, które otrzymujesz od reszty zespołu. Wizualizacja danych jest jak każdy projekt. Nie musisz być ekspertem, aby mieć opinię. Twój zespół będzie najlepszym sposobem, aby dowiedzieć się, czy Twój wykres jest czytelny i zrozumiały. Pamiętaj też, że Twoje najlepsze wykresy będą produktem powstającego projektu. Zacznij od prostych raportów i ulepszaj je z czasem, aby były prostsze i bardziej czytelne. Każda nowa wersja powinna wyjaśniać posiadane informacje. Możesz również ulepszyć raport, dodając informacje bez zwiększania jego złożoności. Otrzymasz znacznie piękniejsze raporty, jeśli przejdziesz przez kilka ulepszonych wersji. Reszta zespołu będzie najlepszym źródłem, które pomoże Ci się tam dostać.

Zrozumienie silosów i uwolnienie danych

Jednym z największych wyzwań dla zespołu zajmującego się analizą danych jest uzyskanie dostępu do wszystkich danych organizacji. Jest to jeden z podstawowych obowiązków kierownika projektu. Będzie ciężko pracować, aby uzyskać dostęp do starych, zakurzonych serwerów baz danych w każdym zakątku Twojej organizacji. Często zdarza się, że duże organizacje mają inną bazę danych dla każdego działu i nie udostępniają informacji. Każdy dział będzie miał własnego analityka danych, kierownika i inżyniera baz danych. Osoby w tym dziale będą wiedziały o swoich wyzwaniach związanych z danymi, ale niewiele wiedzą o innych częściach organizacji. Nazywa się to zwykle silosem danych. Termin silos zapożyczony jest z rolnictwa. Rolnicy zazwyczaj przechowują zboża sezonowe w wysokim, wyodrębnionym budynku. Jeśli kiedykolwiek go widziałeś, wiesz, że silos to wysoki, samodzielny budynek. Każdy silos przechowuje własne ziarno i zwykle nie ma połączenia z innymi silosami. Tak samo jest z danymi

organizacyjnymi. Każdy dział przechowuje fragmenty danych i nie miesza się ich z innymi silosami. Silosy danych istnieją nie bez powodu. Każdy dział może mieć własne wyzwania regulacyjne lub środki bezpieczeństwa. Jedna baza danych może zawierać hasła, a inna może być po prostu prostą listą produktów. Możliwe, że baza haseł jest zaszyfrowana i bezpieczna, a baza produktów jest otwarta i dostępna. Problem z silosami polega na tym, że bardzo utrudniają organizacji całościowy wgląd w dane. Kiedyś pracowałem dla organizacji, która próbowała przekształcić wszystkie swoje dane w jeden scentralizowany klaster. Czuli, że nie uzyskują wystarczającego wglądu w swoje dane. Organizacja przeszła tylko jedną transformację zarządzania danymi i chciała zarządzać sposobem kontrolowania danych. Kiedy w końcu dotarli do swoich danych, zdali sobie sprawę, jak wiele było zamkniętych w silosach, o których nikt nie wiedział. Przez wiele lat każdy dział tworzył własne procesy, schematy i procedury bezpieczeństwa. Chciałem uzyskać wartość z tych danych, ale znajdowały się one na różnych serwerach w całej firmie. Prośbienie wszystkich o udostępnienie swoich danych było trochę jak prośbienie o udostępnienie szczoteczki do zębów. Kierownik projektu był na wielu spotkaniach z bardzo gorącymi dyskusjami. Niestety rozbicie tych silosów jest konieczne, aby stać się organizacją bardziej opartą na danych. Będziesz potrzebować kilku zespołów zajmujących się analizą danych pracujących na jednym scentralizowanym klastrze danych. Będziesz także chciał, aby osoby spoza zespołu analityki danych tworzyły podstawowe raporty i wykresy. Pamiętaj, że dane to kluczowy materiał do znajdowania insightów i tworzenia nowej wiedzy organizacyjnej. Aby być opartym na danych, musisz mieć swobodny dostęp do danych. Teraz, gdy rozumiesz silosy, jak demokratyzować dane, aby były dostępne dla jak największej liczby osób? Musisz upewnić się, że dane nie są już chronionym zasobem przekazywanym między dyrektorami a analitykami danych. Zamiast tego musi to być zasób obejmujący całą firmę, który jest manipulowany i rozumiany przez wszystkich. Jeśli jesteś kierownikiem projektu w zespole zajmującym się analizą danych, pamiętaj o kilku rzeczach:

- Nie lekceważ trudności w uzyskaniu dostępu do silosów danych. To zajmuje dużo czasu, a chcesz zacząć, zanim zespół faktycznie będzie potrzebował ich danych.
- Stworzenie scentralizowanego klastra danych wymaga zmiany organizacyjnej. Jeśli nie masz wpisowego dla kadry kierowniczej, prawdopodobnie nie zrobisz dużych postępów. Prawdopodobnie musisz sprzedać każdemu działowi ideę scentralizowanego klastra danych. Większość działów nie podzieli Twojego zdania. Dochodzą do wniosku, że jeśli nie jest zepsuty, dlaczego mieliby tracić czas na naprawę? Być może będziesz musiał ich zachęcić, wyjaśniając, że będą w stanie tworzyć bardziej złożone raporty lub korzystać z nowszych narzędzi wizualizacyjnych.
- Musisz zapewnić dostęp do raportów swojego zespołu. Łatwiej będzie Ci rozbić silosy, jeśli potrafisz wykazać wartość raportów i analiz w całej firmie. Po pewnym czasie łatwiej będzie wysunąć argument, że współdzielone dane są jak każdy inny współdzielony zasób. Całość będzie lepsza niż suma wszystkich jej części.

Staraj się chronić zespół ds. analizy danych przed spotkaniami dotyczącymi silosu danych. Chcesz, żeby skupili się na eksploracji i odkrywaniu. Skupiasz się na zwiększeniu ich wykorzystania i dostępu.

Tworzenie cyklu spostrzeżeń w zespole Data Science

Wiele organizacji skupia się na monitorowaniu kamieni milowych każdego zespołu. Menedżerowie skupiają się na przestrzeganiu przepisów, a większość ich wysiłków poświęcona jest planowaniu. Mają budżety kwartalne i uważnie je monitorują. Szukają różnic w kosztach lub harmonogramie. Jeśli widzą zmianę, szybko ją ścigają, a następnie zgłaszają ją kierownictwu. Tego typu organizacje są skonstruowane pod kątem monitorowania i zgodności. Jeśli jesteś w tego typu organizacji, pomyśl o swoich spotkaniach. Możliwe, że robisz takie rzeczy, jak planowanie lub przedstawianie planu, koordynacja z innym zespołem, prośba o zwiększenie budżetu lub prośba o przedłużenie

harmonogramu, ponieważ masz opóźnienia. Praca w ten sposób nie jest odpowiednia dla zespołów zajmujących się analizą danych. Pamiętaj, że praca Twojego zespołu ma charakter odkrywczy. Jej członkowie wymyślają pytania, tworzą teorie, a następnie przeprowadzają eksperymenty. Z pewnością są firmy przyzwyczajone do współpracy z naukowcami, takie jak firmy farmaceutyczne czy high-tech. Tego typu firmy od lat prowadzą eksperymenty. Jednak dla większości firm prace eksploracyjne to nowa koncepcja. Posiadanie zespołu zajmującego się analizą danych, który tworzy nową wiedzę, nie wydaje się naturalne. W tych firmach trzeba szczególnie uważać na to, jak pracuje zespół. Pojawi się presja instytucjonalna, aby oddzielić biznes od technologii, a także silny nacisk, aby upewnić się, że zespół ds. Zgodności kieruje zespołem. Zwykle jest to kierownik projektu lub dyrektor. Umieszczenie tych struktur na miejscu może spowolnić tempo odkrywania. Kiedyś pracowałem dla organizacji, która nie pozwalała kierownikowi badań ściśle współpracować z zespołem danych. Uważali, że zadaniem analityków danych jest wymyślanie najlepszych spostrzeżeń, a menedżer biznesowy po prostu widzi pomysły w comiesięcznych prezentacjach. Menedżer biznesowy miał własny budżet, oddzielony od zespołu danych. Menedżer biznesowy nie był zainteresowany znajdowaniem spostrzeżeń. Po prostu upewniła się, że jej zespół mieści się w ich budżecie. Prowadzenie badań w pełnym wymiarze godzin nie mieściło się w ich budżecie. Zespół zajmujący się analizą danych przestał badać, zanim jeszcze się zaczął. Widziałem, jak inna firma próbowała wykorzystać kierownika projektu do monitorowania kamieni milowych zespołu data science. Próbował stworzyć różne sposoby mierzenia postępów zespołu. Stworzył zadania do opracowywania pytań, a następnie zmierzył, jak dobrze zespół wykonał te zadania. To nie działało dobrze, ponieważ większość pytań prowadzi do większej liczby pytań. Kierownik projektu nie był zadowolony, gdy kamienie milowe zespołu wciąż się przesunęły. Jego motywacją było jak najszybsze „uzupełnienie” pytań, co jest przeciwieństwem tego, czego oczekujesz od zespołu zajmującego się badaniem danych. Kiedy jesteś w zespole zajmującym się analizą danych, staraj się zdawać sobie sprawę z tych presji instytucjonalnych. Większość organizacji ma trudności z zaakceptowaniem zespołu zajmującego się analizą danych, którego nie można łatwo zmierzyć ani kontrolować. Trudno jest wyznaczyć cele lub stworzyć zwrot z inwestycji. Ciężko pracuj, aby upewnić się, że Twój zespół nie zostanie wciągnięty w te zgodne struktury. Zamiast tego upewnij się, że zespół tworzy pętlę sprzężenia zwrotnego. Wszyscy powinni współpracować, aby kwestionować, badać i uczyć się. Każda osoba w zespole ds. analizy danych ma swój własny obszar zainteresowania, ale nadal współpracują ze sobą w wąskich pętlach sprzężenia zwrotnego. Na przykład kierownik projektu uczestniczy, gdy kierownik badania i analityk danych eksplorują dane, a analityk danych może dać kierownikowi projektu dobre sugestie dotyczące uzyskania dostępu do danych innego zespołu. Drużyna zawsze radzi sobie lepiej, gdy wszyscy biorą udział. Kiedy eksplorujesz, im więcej oczu masz na problemie, tym większe prawdopodobieństwo, że uzyskasz świetne wglądy. Upewnij się też, że Twój zespół ds. analizy danych ma wsparcie na poziomie wykonawczym. Bez tego prawie na pewno zostaniesz wciągnięty z powrotem do wspólnych strategii kontroli. Nie będzie to łatwe, ale jeśli masz wsparcie na poziomie kierowniczym i możesz tworzyć wąskie pętle zwrotne w zespole, znacznie łatwiej będzie ci dokonywać odkryć. Nie denerwuj się, jeśli wprowadzenie tych zmian organizacyjnych zajmie dużo czasu. Pierwszym krokiem jest zrozumienie celu nauki o danych i tego, jak mieści się ona poza tymi od dawna działającymi strukturami organizacyjnymi.

PODSUMOWANIE

Zbadałeś obszary odpowiedzialności każdego członka zespołu. Potem zobaczyłeś, gdzie nakładają się na siebie te obszary odpowiedzialności. Dowiedziałeś się również trochę o tym, jak prezentować swoje dane (wizualizacja danych) oraz o różnicy między raportami wewnętrznymi i zewnętrznymi. Następnie dowiedziałeś się o silosach danych i ich wpływie na zdolność Twojego zespołu do zbierania danych. Wtedy otrzymałeś kilka wskazówek, jak zdemokratyzować swoje dane. Na koniec dowiedziałeś się, jak

ważne jest zapewnianie przejrzystości zespołowi ds. analizy danych i jak to pomaga w zaangażowaniu kadry kierowniczej.